Solutions

March 20, 2016 STA 5327

1. The information from the study is as follows:

$$X_1 = x_1, X_2 = x_2, \ldots, X_r = x_r, X_{r+1} > T, X_{r+2} > T, \ldots, X_n > T.$$

The likelihood based on the observations is

$$L(\theta) = \theta e^{-\theta x_1} \theta e^{-\theta x_2} \ldots \theta e^{-\theta x_r} e^{-\theta T} e^{-\theta T} \ldots e^{-\theta T}$$

where the first r terms appear from the density of the exponential distribution and the next $n-r$ terms appear from the of the exponential distribution. Hence

$$\begin{aligned} L(\theta) &= \theta^r \exp\bigg(-\theta \sum_{i=1}^r x_i\bigg) \exp\{-(n-r)\theta T\} \\ &= \theta^2 \exp\bigg[-\theta\bigg\{\sum_{i=1}^r x_i + (n-r)T\bigg\}\bigg] \end{aligned}$$

and the log-likelihood is

$$l(\theta) = r \log \theta - \theta\bigg[\sum_{i=1}^r x_i + (n-r)T\bigg].$$

It is easy to see that the maximum of $l(\theta)$ occurs at the stationary point obtained as a solution of

$$\frac{\partial l}{\partial \theta} = 0 \Leftrightarrow \frac{r}{\theta} - \bigg[\sum_{i=1}^r x_i + (n-r)T\bigg] = 0$$

leading to

$$\hat{\theta} = \frac{r}{\sum_{i=1}^r x_i + (n-r)T}.$$

Emphasize in this problem that writing the likelihood is the main step

2. There can be many choices for an unbiased estimator $T(X)$. One such example can be obtained as letting $T(2) = 4$ and $T(x) = 0$ for $x \neq 4$. Then

$$\mathbb{E}[T(X)] = 4 \times \theta/4 = \theta.$$

Since only one data point is observed and X can take only 5 values, it is sufficient to find the MLE in these 5 cases.

$$\begin{aligned}
\hat{\theta}(-2) &= \text{arg max} \quad (1-\theta)/4 = 0 \\
\hat{\theta}(-1) &= \text{arg max} \quad \theta/12 = 1 \\
\hat{\theta}(0) &= \text{arg max} \quad 1/2 = \quad \text{any number in } (0,1) \\
\hat{\theta}(1) &= \text{arg max} \quad (3-\theta)/12 = 0 \\
\hat{\theta}(2) &= \text{arg max} \quad \theta/4 = 1.
\end{aligned}$$

Since $\hat{\theta}(0)$ can take multiple values, MLE is not unique when the datapoint $X = 0$ is observed. To see that all the MLEs are biased, we try to solve the equation

$$\mathbb{E}[\hat{\theta}] = 1 \times \theta/12 + 1 \times \theta/4 + \hat{\theta}(0) \times 1/2 = \theta$$

for $\hat{\theta}(0)$. Clearly, this amounts to having $\hat{\theta}(0) = 4\theta/3$ which is impossible since $\hat{\theta}(0)$ must not depend on θ. Hence all the MLEs are biased.

3. (a) Observe that

$$F(x,y) = 1 - \mathbb{P}(X_1 \le x) - \mathbb{P}(Y_1 \le y) + \mathbb{P}(X_1 \le x, Y_1 \le y).$$

Note that $F(x,y)$ is differentiable when $x \neq y$ but not when $x = y$. When $x \neq y$, (X_1, Y_1) has density obtained using $\frac{\partial^2 F(x,y)}{\partial x \partial y}$ as

$$f_\theta(x,y) = \begin{cases} (\theta+1)(1-x)^\theta, & x > y \\ (\theta+1)(1-y)^\theta, & x < y \end{cases}$$

Let $X = X_1$ and $Y = Y_1$, we have

$$\begin{aligned}
\mathbb{P}(X > t, Y > t, X \neq Y) &= 2\mathbb{P}(X > t, Y > t, X > Y) \\
&= 2(\theta+1)\int_t^1 \int_t^x (1-x)^\theta dy dx \\
&= \frac{2(1-t)^{\theta+2}}{\theta+2}.
\end{aligned}$$

Also

$$\mathbb{P}(X > t, Y > t) = (1-t)^{\theta+2}.$$

Hence

$$\begin{aligned}
\mathbb{P}(X > t, X = Y) &= \mathbb{P}(X > t, Y > t) - \mathbb{P}(X > t, Y > t, X \neq Y) \\
&= \frac{\theta(1-t)^{\theta+1}}{\theta+2}
\end{aligned}$$

which means (X,Y) has density $\theta(1-t)^{\theta+1}$ on the line $x=y$. Then the probability density (strictly speaking, a distribution since its not absolutely continuous with respect to the Lebesgue measure) of (X,Y) is

$$f_\theta(x,y) = \begin{cases} (\theta+1)(1-x)^\theta, & x > y \\ (\theta+1)(1-y)^\theta, & x < y \\ \theta(1-x)^{\theta+1}, & x = y. \end{cases}$$

(b) Let T be the number of (X_i,Y_i)'s with $X_i = Y_i$ and $V_i := \max\{X_i, Y_i\}$, the likelihood can be re-written as

$$L(\theta) = (\theta+1)^{(n-T)}\theta^T \prod_{i=1}^n (1-V_i)^\theta \prod_{i:X_i=Y_i} (1-V_i).$$

Letting $\frac{\partial l(\theta)}{\partial \theta} = 0$, we can find that the unique solution in the parameter space for θ is

$$\hat\theta = \frac{\sqrt{(n-W)^2 + 4WT} + (n-W)}{2W},$$

where $W = -\sum_{i=1}^n \log(1-V_i)$. Since

$$\frac{\partial^2 l(\theta)}{\partial \theta^2} = -\frac{n-T}{(\theta+1)^2} - \frac{T}{\theta^2} < 0$$

$\hat\theta$ is the MLE of θ.

4. (a) It is enough to show that the negative log-likelihood function $l(\theta) = -\log\prod_{i=1}^n f(x_i \mid \theta)$ is a strictly convex function of $\theta \in \mathbb{R}^p$. Since its the sum of the negative likelihoods for each X_i, and a sum of strictly convex functions is strictly convex, it is enough to consider a single observation $n=1$. To show that the negative log likelihood is strictly convex it is enough to show that its Hessian, the matrix $H_{ij} = \frac{\partial^2}{\partial\theta_i \partial\theta_j} l(\theta)$ is positive definite everywhere (except possibly at $\hat\theta = x$). Let's

compute the necessary derivatives:

$$\begin{aligned}
l(\theta) &= -\log c_\alpha + \left(\sum_{i=1}^p (x_i-\theta_i)^2\right)^{\alpha/2} \\
\frac{\partial l(\theta)}{\partial \theta_k} &= -\alpha\left(\sum_{i=1}^p (x_i-\theta_i)^2\right)^{(\alpha-2)/2} (x_k-\theta_k) \\
\frac{\partial^2 l(\theta)}{\partial \theta_k^2} &= \alpha\left(\sum_{i=1}^p (x_i-\theta_i)^2\right)^{(\alpha-2)/2} + \alpha(\alpha-2)\left(\sum_{i=1}^p (x_i-\theta_i)^2\right)^{(\alpha-4)/2} (x_k-\theta_k)^2 \\
&= \alpha\left(\sum_{i=1}^p (x_i-\theta_i)^2\right)^{(\alpha-4)/2} \left[\sum_{i=1}^p (x_i-\theta_i)^2 + (\alpha-2)(x_k-\theta_k)^2\right] \\
\frac{\partial^2 l(\theta)}{\partial \theta_i \partial \theta_k} &= \alpha\left(\sum_{i=1}^p (x_i-\theta_i)^2\right)^{(\alpha-4)/2} (\alpha-2)(x_i-\theta_i)(x_k-\theta_k).
\end{aligned}$$

The Hessian $H = \alpha \|x-\theta\|^{\alpha-4} A$ is the product of a constant positive factor $\alpha \|x-\theta\|^{\alpha-4}$ and a matrix A whose on and off diagnonal entries are:

$$\begin{aligned}
A_{kk} &= \sum_{i=1}^p (x_i-\theta_i)^2 + (\alpha-2)(x_k-\theta_k)^2 \\
A_{kj} &= (\alpha-2)(x_k-\theta_k)(x_j-\theta_j).
\end{aligned}$$

Introducing the notation $\Delta = (x-\theta) \in \mathbb{R}^p$ for the vector with components $\Delta_i = (x_i-\theta_i)$ and I_p for the $p \times p$ identity matrix, we can write A in the form

$$A = \|\Delta\|^2 I_p + (\alpha-2)\Delta\Delta'.$$

The matrix A satisfies $A\Delta = \lambda\Delta$ with eigenvalue $\lambda = \|\Delta\|^2(\alpha-1)$, strictly positive since $\alpha > 1$ (except at $\Delta = 0$, i.e., $\theta = \hat{\theta} = x$, which is okay). The other eigenvectors are orthogonal to Δ, all with eigen values $\lambda' = \|\Delta\|^2$, which are also strictly positive. Thus A is a positive definite matrix and so is the Hessian, $H = \alpha \|\Delta\|^{\alpha-4} A$.

(b) First consider the case where $\alpha = 1$ in dimension $p = 1$, with $n = 2m$ even. Without loss of generality, order the data $x_1 \le x_2 \le \cdots \le x_n$. The log likelihood function is given by

$$l(\theta) = n \log c(\alpha) - \sum |x_i - \theta|,$$

continuous function whose derivative does not exist at the data points $\theta \in$

$\{x_1, \ldots, x_n\}$, and which elsewhere satisfies

$$
\begin{aligned}
\frac{d}{d\theta} l(\theta) &= -\sum \frac{d}{d\theta} |x_i - \theta| \\
&= \left[\sum_{i:x_i<\theta} (-1) \right] + \left[\sum_{i:x_i>\theta} (+1) \right].
\end{aligned}
$$

Note that the derivative of $|x - \theta|$ is -1 on the interval $\theta \in (-\infty, x)$, 1 on the interval $\theta \in (x, \infty)$, and undefined at the point $\theta = x$). Thus $l(\theta)$ is increasing when $\theta < x_m$, when more than half the $\{x_i\}$ exceed θ, and is decreasing when $\theta > x_{m+1}$, when fewer than half the $\{x_i\}$ exceed θ. In the interval $x_m < \theta < x_{m+1}$, the derivative is zero, so $l(\theta)$ is constant there and equal to its maximum value. In case $n = 2m - 1$ is odd, the same argument shows that $l(\theta)$ achieves a unique maximum at the median x_m. In dimension $p > 2$ a similar argument holds, only $\hat{\theta}$ now should be any value in the "median rectangle" (or "median block") where each of its components is a median of the corresponding components of the x_i's.